Engineering Novel S-Glycosidase Activity into Extremo-Adapted β-Glucosidase by Rational Design

Nouarh Almulhim[1], Nicholas R. Moody[1] and Francesca Paradisi[1,2*]

[1]University of Nottingham, School of Chemistry, University Park, Nottingham NG7 2RD, United Kingdom

[2]Department of Chemistry and Biochemistry University of Bern, Freiestrasse 3, 3012, Bern, Switzerland.

*Corresponding author. Department of Chemistry and Biochemistry University of Bern, Freiestrasse 3, 3012, Bern, Switzerland. Tel: +41 31 631 40 02 E-mail : francesca.paradisi@dcb.unibe.ch.

Nouarh Almulhim ORCID iD: 0000-0002-0845-4918

Nicholas R. Moody ORCID iD: 0000-0002-0662-686X

Francesca Paradisi ORCID iD: 0000-0003-1938-262X

ABSTRACT

The breakdown of sulphur glycosidic bonds in thioglycosides can produce isothiocyanate, a chemoprotective agent linked to the prevention of cancers, however only a handful of enzymes have been identified that are known to catalyse this reaction. Structural studies of the myrosinase enzyme, which is capable of hydrolysing the thioglycosidic bond, has identified residues that may play important roles in sulphur bond specific activity. Using rational design, two extremo-adapted β-glycosidases from the species *Thermus nonproteolyticus* (*Tno*GH1) and *Halothermothrix orenii* (*Hor*GH1) were engineered towards thioglycoside substrates. Twelve variants, six for *Tno*GH1and six for *Hor*GH1, were assayed for activity. Remarkable enhancement of the specificity (k_{cat}/K_M) of *Tno*GH1 and *Hor*GH1 towards β-thioglycoside was observed in the single mutants *Tno*GH1-V287R (2500 $M^{-1}s^{-1}$) and *Hor*GH1-M229R, (13260 $M^{-1}s^{-1}$) which showed a 3-fold increase with no loss in turnover rate when compared to the wild type enzymes . Thus, the role of arginine is key to induce β-thioglycosidase activity. Thorough kinetic investigation of the different mutants has shed light on the mechanism of β-glycosidases when acting on the native substrate.

KEY POINTS

Key residues were identified in the active site of *Brevicoryne brassicae* myrosinase.

Rationally designed mutations were introduced into two extremo-adapted β-glycosidases.

β-glycosidases mutants exhibited improved activity against thioglycosidic bonds.

The mutation to arginine in the active site yielded the best variant.

KEYWORDS

Enzyme engineering, Site directed mutagenesis, glycoside hydrolase, *Thermus nonproteolyticus*, *Halothermothrix orenii*, myrosinase

INTRODUCTION

The glycoside hydrolase family 1 of enzymes (GH1; EC 3.2.1.21) is characterised by the ability to catalyse the hydrolysis of glycoside linkages in a variety of sugars (β-glucosides) (Davies and Henrissat 1995; Park et al. 2017). In terms of structure, it has been demonstrated that GH1 are $(\beta/\alpha)_8$ barrel folded enzymes, which consist of eight twisted, parallel β-strands, located in the internal part of the protein, surrounded by eight α-helixes in the external part. The C-terminal (in the β strand 8) of all known $(\beta/\alpha)_8$ barrel proteins hosts the active site residues within the β → α loop (Henn-Sax et al. 2001; Silverman et al. 2001). The hydrolytic ability of GH1 is dependent on two critical glutamic acid residues (Fig. 1), E164 (*Thermus nonproteolyticus* glycoside hydrolase, *Tno*GH1 numbering), located in the $T^{161}LNEP^{165}$ motif (β-strands 4), is the acid catalyst, and a second one, E338, in the $I^{336}TENG^{340}$ motif (β-strands 7) is the nucleophile (Wang et al. 2003) (Fig. 1a). E164 plays an important role in the formation of the intermediate (enzyme-saccharide) of classical glycosidases as an activator of the glycosidic oxygen.

Thioglycosides are among the most stable glycosidic molecules. In these structures, the glycoside is bridged to the aglycon moiety by a sulphur bond. The breakdown of S- glycosidic bonds in glucosinolates (GSL) can release molecules with activity against pests and herbivores (part of the plant defence mechanism) and isothiocyanate, a chemoprotective agent linked to the prevention of cancers (Dufour et al. 2015; Halkier and

Gershenzon 2006; Rakariyatham et al. 2005 ; Samec et al. 2016; Winde and Wittstock 2011).

Myrosinases (EC 3.2.3.147) are unique members of the GH1 family able to hydrolyse thioglycosides. Unlike β -glycosidases which are ubiquitous, myrosinases have been identified only in a handful of species such as *Sinapis alba* (Burmeister et al. 1997), *Brevicoryne brassicae* (Jones et al. 2002), *Verticillium longisporum* (Witzel et al. 2015)*, Arabidopsis thaliana,* and *Brassica napus* (Nong et al. 2010). The active site of *Sinapis alba* myrosinase (*Sa*MYR), a plant species, differs from that of classical β -glycosidases, as it lacks the catalytic glutamic acid residue in the $T^{184}INQL^{188}$ motif (equivalent to $T^{161}LNEP^{165}$ in *Tno*GH1) (Bourderioux et al. 2005), while it maintains the second one (E409 in motif $T^{408}ENG^{411}$) (Burmeister et al. 1997). (Fig. 2a). In addition, *Sa*MYR requires ascorbic acid as a cofactor to catalyses the hydrolysis of thioglycosidic substrates.

A myrosinase from the cabbage aphid *Brevicoryne brassicae (Bb*MYR) on the other hand, relies on the typical catalytic acid/base system found in β-glycosidases (E167 and E374) (Fig. 2b) and it more closely aligns with classic β-glycosidases than *Sa*MYR in term of structure and amino acid residues present in the active site. *Bb*MYR has however unique structural features, not observed in either β-glycosidases or other myrosinases. K173 and R312 (Fig. 3) play a critical role in the hydrolysis of GSL, as they are directly involved in its recognition (Jones et al. 2002). Y180 may also play a role due to its proximity to the thioglycosidic linkage in the substrate (Husebye et al. 2005). Figure S1 shows full sequence alignment of *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1), *Halothermothrix orenii (Hor*GH1*),* cabbage aphid *Brevicoryne brassicae (Bb*MYR*), and Sinapis alba* myrosinase (*Sa*MYR).

A thermophilic GH1 from the extremophilic bacterium *Thermus nonproteolyticus* (*Tno*GH1) and a halotolerant GH1 isolated from *Halothermothrix orenii* (*Hor*GH1) have been previously described in the literature (He et al. 2001; Kori et al. 2011). Enzymes from extremophile organisms have significantly higher tolerance than the mesophilic counterparts to temperature and/or pHs, making them attractive for industrial applications (Yin et al. 2015), however, no known examples have been reported of extremophilic GH1 with myrosinase activity. Engineering extremo-adapted GH1 to broaden their substrate scope towards β-thioglycosidase activity could significantly increase their potential applications in an industrial setting.

Here we report how rational design aided the introduction of amino acid mutations by mapping the *Bb*MYR active site onto the extremophilic *Tno*GH1 enzyme first, and then onto *Hor*GH to further confirm the key role played by selected residues in the recognition and hydrolysis of thioglycosides. In *Tno*GH1, the mutations L171K, V287R, and H178Y were introduced as single and double mutants (generating 3 additional variants with all possible permutations). The equivalent single and double mutants at positions E173K, M299R, and H180Y were then introduced into *Hor*GH1. The kinetic properties of all variants and wild type enzymes with test substrates β-D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) have been carried out and the role of arginine (*Tno*GH1-V287R and *Hor*GH1-M299R) has been found to be pivotal as a marked improvement of activity towards the β-thioglucoside substrate has been observed in both mutant proteins. These are the first examples of extremophilic GH1s in which myrosinase activity has been introduced.

MATERIALS AND METHODS

Reagents and bacterial strains

Substrates 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) were purchased from Carbosynth. *Escherichia coli* bacterial strains and QuikChange II Site-Directed Mutagenesis Kit were from Agilent Technologies. The plasmid miniprep kit was from Macherey-Nagel. Growth media and assay components were procured from Fisher Scientific. All other chemicals were purchased from Sigma.

DNA preparation and site-directed mutagenesis

Mutations were introduced using the QuikChange Site-Directed Mutagenesis kit. pCH93b hosting the ds-DNA of *Tno*GH1 (GenBank accession number AF225213) and pET45b hosting the ds-DNA of *Hor*GH1(GenBank accession number WP_012636460) respectively were used as templates in the PCR reactions (Heckmann 2017). The mutagenic primers were designed using QuikChange Primer Design Program (www.agilent.com/genomics/qcpd). Primers are summarised in Table S1.

Protein expression and purification

For protein expression, BL21(DE3) *E. coli* strain was used. Chemically competent *E. coli* cells were transformed with each plasmid. A 300 mL LB flask was inoculated with starter culture and grown at 37 °C (200 rpm) to an OD_{600} of ~ 0.6, prior to induction with 1 mM IPTG. Induced cultures were then incubated at 37 °C overnight. Cells were harvested at 3,500 × *g*, 4°C, 20 min. The pellet was resuspended in loading buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 10 mM imidazole) and lysed by sonication for 20 minutes on ice (1 min on, 30 sec off; 20 cycles). The soluble fraction was decanted following centrifugation at 22,800 × *g*, 4 °C for 1 hour, and filtered with a 0.45 µm filter. Filtered supernatant was loaded onto a 1 mL HisTrap FF crude® column, using an AKTA™ Start. The column was washed with eight column volumes of loading buffer, followed by fifteen column volumes of loading buffer with 10 % elution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 300 mM imidazole). The protein was then eluted with eight column volumes of 100% elution buffer. Pure fractions were pooled and dialysed for 20 hours at room temperature with dialysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl), with one buffer exchange after 2h.

Enzyme quantification

The concentration of the purified enzymes was determined by absorbance at 280 nm. The extinction coefficient ε was estimated using the EXPasy ProtParam tool (Gasteiger et al. 2005) (Table S2). Proteins were analysed with a 12% SDS PAGE, by staining with InstantBlue (Fig. S2).

Kinetics assays

Enzyme activity was measured spectroscopically in triplicate by monitoring the change in absorbance at 420 nm of the p-nitrophenol or p-nitrothiophenol. The extinction coefficients for the products was determined using a calibration curve (Fig. S3).

Assays were conducted in 200 µL at 50°C for *Tno*GH1 and at 25°C for *Hor*GH1. A typical reaction mixture contained 100 mM HEPES buffer pH 7.5, 500 mM sodium chloride. pNT-Glc was dissolved in 30% DMSO, the concentration of DMSO was controlled at 9 % across pNT-Glc assays. All assay components were filtered with a

0.45 µm filter prior to use. Assays were initiated with the addition of enzyme. Primary non-linear regression plots are described in Fig S5, S6, S7 and S8.

Data analysis

In silico modelling of crystal structures was performed using the UCSF Chimera software (Pettersen et al. 2004). Sequence alignments were determined with ENDscript server software (Robert and Gouet 2014). Similarities and identities, including homology modelling, between sequences were calculated using EMBOSS software (Rice et al. 2000). Kinetic parameters were evaluated by nonlinear regression analysis in Igor Pro (Babonneau 2010). Bar graphs were produced in GraphPad Prism (Swift 1997). The primary plots were analysed using equation 1. For consensus analysis of amino acids, β-glycosidase sequences were taken from the Pfam protein family's database (El-Gebail et al. 2019), ~4000 sequences of representative proteome were used for amino acid analysis. The representative proteome at 15% comembership threshold, as defined by Pfam which is an even sampling of the sequences of the glycosyl hydrolase family PF00232 (Chen et al. 2011), was aligned with the wild type sequences of the β-glycosidase used in this study with the ClustalW tool with the MEGA X software (Kumar et al. 2018). Sequence logo analysis was performed with WebLogo (Crooks et al. 2004).

$$\frac{v_i}{[E]_T} = \frac{k_{cat}^{app}[S]}{K_M^{app} + [S]} \qquad \text{(Equation 1)}$$

RESULTS

Computational study of *Tno*GH1, *Hor*GH1, *Bb*MYR, and *Sa*MYR

A comparative study between the amino acid sequences of the four glycosyl hydrolases (*Tno*GH1, *Hor*GH1, *Bb*MYR, and *Sa*MYR) was carried out (Fig. S1). Both *Tno*GH1 and *Hor*GH1, as expected, present a greater sequence similarity to the *Bb*MYR than *Sa*MYR. *Tno*GH1 and *Bb*MYR show a 53.6% similarity (34.3% identity), greater than that observed between *Tno*GH1 and *Sa*MYR (46.3% similarity and 30.0% identity). *Hor*GH1 has a 58.1 % similarity (37.2 % identity) with *Bb*MYR and a 51.9% (33.1 % identity) with *Sa*MYR. The active site is highly conserved across all enzymes.

*Tno*GH1 and *Hor*GH1mutants construction

Three residues in the amino acid sequence of *Bb*MYR - K173, R312, and Y180, were identified to be important in the hydrolysis of thioglucosides (Jones et al. 2002; Husebye et al. 2005). The sequence alignment between *Bb*MYR and *Tno*GH1 (Fig. 3) reveals that *Tno*GH1 presents a leucine, a valine, and a histidine at the equivalent positions (L171, V287, and H178 - *Tno*GH1 numbering). Likewise, in *Hor*GH1 (Fig. 3) three residues (E173, M299 and H180) were selected to be replaced by K, R, and Y, respectively. Three single mutants were generated in both the thermotolerant enzyme *Tno*GH1 (L171K, V287R, and H178Y) and the halotolerant *Hor*GH1 (E173K, M299R, and H180Y). Double mutant permutations were then created (L171K/V287R, L171K/H178Y, and V287R/ H178Y in *Tno*GH1, and E173K/ M299R, E173K/ H180Y, and M299R/ H180Y in *Hor*GH1) to elucidate any synergistic effect among these amino acids. All mutants were expressed and purified (Fig. S2).

Wild type *Tno*GH1 and *Hor*GH1 substrate scope

*Bb*MYR activity against sinigrin (Fig. 4), the native substrate, was reported with k_{cat} = 36 s^{-1}, K_M = 0.41 mM (Pontoppidan et al. 2001). Both *Tno*GH1 and *Hor*GH1 exhibited some activity against pNT-Glc, used here as

substrate mimic. All determined kinetic parameters are summarised in Table 1 and Table 2. For *Tno*GH1, k_{cat} of 1.34 s^{-1}, K_M of 1.43 mM and k_{cat}/K_M 940 $M^{-1}s^{-1}$ were observed (Fig. S4). For *Hor*GH1, *k*cat of 37.20 s^{-1}, K_M of 4.10 mM and k_{cat}/K_M 9083 $M^{-1}s^{-1}$ were measured (Fig. S4), identifying the latter as the better catalyst towards pNT-Glc (k_{cat} of ~28-fold higher than *Tno*GH1).

***Tno*GH1 and *Hor*GH1 single mutant kinetic studies (pNT-Glc)**

When compared to the wild type, the *Tno*GH1- V287R single mutant shows the greatest increase in specificity, k_{cat}/K_M, and a retention of turnover rate, k_{cat} (Table 1) with pNT-Glc.

The mutant *Tno*GH1-L171K shows a near retention in specificity (1200 $M^{-1}s^{-1}$) and turnover rate (1.10 s^{-1}), while *Tno*GH1-H187Y shows a 3-fold loss in specificity (300 $M^{-1}s^{-1}$) and a 10-fold loss in turnover rate (0.15 s^{-1}).

*Tno*GH1-V287R shows an improvement in specificity with no loss of turnover rate. The combined improvement in k_{cat}/K_M and retention of k_{cat} identifies *Tno*GH1-V287R as the best mutant for practical application. The kinetic parameters of *Tno*GH1 are summarised in Table 1.

A similar pattern is observed in the *Hor*GH1 mutants. *Hor*GH1-M299R shows over 30% increase in specificity and a retention of turnover number (Table 2). While little change is observed in *Tno*GH1-L171K, *Hor*GH1-E173K shows a 3-folld loss in specificity and a 2-fold loss in turnover rate. The *Hor*GH1-H180Y mutant again shows a 3-fold decrease in specificity and a 4-fold decrease in turnover compared to the wild type, similar to the change observed in the respective mutant in *Tno*GH1. All kinetic parameters of *Hor*GH1are summarised in Table 2.

***Tno*GH1 and *Hor*GH1 double mutant kinetic studies (pNT-Glc)**

All the *Tno*GH1 double mutants show a lower turnover rate with the target pNT-Glc compared to the wild type and, with the exception of the *Tno*GH1-V287R/H178Y, to the single mutants. In particular, *Tno*GH1-L171K/V287R shows a 2-fold increase in specificity (1820 M^{-1} s^{-1}) but a 3-fold decrease in turnover rate. The *Tno*GH1-L171K/ H178Y mutant shows a 3-fold decrease in specificity (380 M^{-1} s^{-1}) as well as a 10-fold decrease in turnover rate. The *Tno*GH1-V287R/H178Y mutants shows a 2.5-fold increase in specificity and a 1/3-fold decrease in turnover rate. *Tno*GH1-V287R/H178Y is the only variant in this series with a specificity (2420 M^{-1} s^{-1}) comparable to that observed in the single mutant *Tno*GH1-V287R (2500 M^{-1} s^{-1}).

With *Hor*GH1 double mutants, the *Hor*GH1-E173K/ M299R shows a 3-fold decrease in specificity and a 4-fold decrease in turnover rate compared to the wild type (Table 2). The *Hor*GH1-E173K/H180Y mutant shows an 8-fold decrease in specificity and an 8-fold decrease in turnover. The *Hor*GH1-M299R/H180Y mutant shows a near 3-fold decrease specificity and a 4-fold decrease in turnover rate, similar to that seen in *Hor*GH1-E173K/ M299R.

As with single mutants, the pattern observed in *Tno*GH1 double mutants is closely mapped in the *Hor*GH1 variants. All double mutants show a lower turnover rate compared to that of the single mutants or wild type. Double mutants containing the arginine mutation (*Tno*GH1-V287R and *Hor*GH1-M299R) show an increase in specificity. Kinetic parameters summarised in Table 1 and 2.

***Tno*GH1 and *Hor*GH1 variants: substrate analysis**

The *Tno*GH1-V287R mutant shows a significant shift in specificity towards pNT-Glc which corresponds to 30% loss in specificity towards pNP-Glc without any loss in turnover rate (Table 1). In the analogous *Hor*GH1 mutant, this observation is pronounced, as there is a 35-fold loss in specificity towards the O-glycosidic substrate, as well as a 2-fold decrease in turnover number (Table 2). When comparing substrates, kinetic parameters for both enzymes appear to diverge to some extent in this case.

With pNT-Glc, *Tno*GH1-L171K shows a retention of kinetic parameters however with pNP-Glc, there is a 2.5-fold loss in specificity and a retention of turnover (Table 1). In *Hor*GH1-E173K a decrease in kinetic parameters is observed with pNT-Glc, with pNP-Glc this is more pronounced, a 10-fold decrease in specificity is observed with a 2-fold decrease in turnover rate (Table 2). This mutation results in little perturbation in turnover rate for both substrates in *Tno*GH1, however a decrease in parameters is observed for both substrates in *Hor*GH1 (Table 1).

*Tno*GH1-H178Y mutant shows the lowest single mutant turnover rate with pNP-Glc, representing a 2-fold loss (Table 1). Similarly, *Hor*GH1-H180Y also shows a 2-fold decrease in specificity and turnover rate with the same substrate (Table 2). Kinetics parameters change in the same direction for both substrates in both enzymes on mutation at this position (Table 1 and Table 2).

Double mutants also show a complex relationship with respect to the native substrate. *Tno*GH1-L171K/V287R shows a 3-fold loss in specificity and a 1/3-loss in turnover rate (Table 1) compared to the wild type. Whereas with pNT-Glc, this variant shows an increase in specificity (Table 1). The *Tno*GH1-L171K/H178Y mutant shows similar result with pNP-Glc as with pNT-Glc compared to the wild type, a 3-fold loss in specificity is observed in both specificity and turnover rate for pNP-Glc (Table 1). *Tno*GH1-V287R/H178Y shows a similar specificity towards pNP-Glc as the wild type enzyme, but a 1/3 - loss in turnover rate compared to wild type, while with pNT-Glc a gain in specificity is observed compared to the wild type. (Table 1).

When *Hor*GH1 double mutants are compared to the wild type with pNP-Glc substrate, all mutants show a loss in specificity and turnover rate. This is similar to the pattern observed with these mutants and the pNT-Glc. Unlike *Tno*GH1, the double mutants on *Hor*GH1 show the same change in kinetic parameters for both substrates. (Table 2).

DISCUSSION

Significantly enhanced activity towards a thiosaccharide substrate, pNT-Glc, was introduced in *Tno*GH1 and *Hor*GH1 through rational design. The *Tno*GH1-V287R and *Hor*GH1-M299R variants yielded the greatest increase in specificity towards pNT-Glc while retaining similar turnover number to the wildtype. It has been shown previously that R312 play a critical role in aphid myrosinase for sulfur recognition (Jones et al. 2002). In β-glycosidases, the amino acids at the equivalent position are hydrophobic in nature (valine in *Tno* and methionine in *Hor*). A mutation to arginine introduces a guanidinium side chain in the active site capable of hydrogen bonding to the thioglycosidic bond in the substrate. R312 may have a potential interaction that may stabilize the transition state resulting in the observed increase in k_{cat}/K_M. *Tno*GH1-L171K and *Hor*GH1-E173K, it also adds a positively charged side chain into the active site, however an increase in the k_{cat}/K_M is not observed. The crystal structure of

*Bb*MYR indicates K173 pointing away from the active site, possibly reducing its direct involvement in substrate binding.

The introduction of a tyrosine in both *Tno*GH1 and *Hor*GH1 to replace a histidine results in a dramatic decrease in k_{cat}/ K_M towards both pNP-Glc and pNT-Glc substrates compared to the wild types. Y180 (*Bb*MYR numbering) had been suggested to have a possible catalytic role due to the proximity of the sidechain to the thiosidic linkage in glucosinolates. However, in both extremophilic enzymes, the histidine displayed at that position appears to be highly conserved among β-glycosidases (Fig. 5), suggesting that a mutation at this position is poorly tolerated. Specifically, tyrosine while not forbidden, has an incidence of less than 3% in the data set. The others two targeted positions (L171 and V287, *Tno*GH1 numbering) are not as highly conserved (Fig. S9)

A more complex relationship is observed when the double mutants are compared to the wild type and single mutants. From the single mutant results, we observe a correlation between the introduction of polar residues in the active site of *Tno*GH1 and the specificity for the pNT-Glc substrate. This is supported by the change in specificity observed between the *Tno*GH1-H178Y mutant and the *Tno*GH1-V287R/H178Y double mutant, where the latter has a much-improved specificity when compared to the former. Likewise, when *Tno*GH1-V287R is combined with *Tno*GH1-L171K, the double mutant enzyme has a higher specificity than the *Tno*GH1-L171K but lower specificity than the *Tno*GH1-V287R variant.

We can also see that a similar effect was observed in *H*orGH1 specificity of pNT-Glc substrate with the *Hor*GH1-M299R/H180Y double mutant. Unlike *Tno*GH1, double mutant *Hor*GH1-E173K/M299R shows similar specificity to the *Hor*GH1-E173K single mutant. We see a larger K_M (pNT-Glc) in the double mutant when compared to the *Hor*GH1-M299R single mutant, and a large reduction in k_{cat} compared to both single mutants. This would suggest that the same effect in the *Tno*GH1 mutants may be playing a role in the *Hor*GH1 mutants.

The *Hor*GH1 and *Tno*GH1 mutants all exhibited a decrease in turnover rate and specificity with the native substrate pNP-Glc (Table 1 and 2). This indicates that the increase in specificity for pNT-Glc is at the expense of the native substrate. The increase observed in k_{cat}/ K_M (pNT-Glc) induced by the introduction of the arginine residues, is not observed with pNP-Glc. Considering the catalytic efficiency of the *Hor*GH1-M299R, this mutant has a ratio of k_{cat}/ K_M of glycoside to thioglycoside of 3:1, compared to the *Hor*GH1 wild type which has a ratio of 11:1.

In this study, enhanced β-thioglycosidase activity was introduced by rational design in the extremophilic β-glycosidases *Tno*GH1 and *Hor*GH1 by in-silico modelling of the *B. Brassicae* myrosinase. A three-fold increase in specificity for the thioglycosidic substrate with no loss in turnover number was observed by replacing of hydrophobic residues of both enzymes by arginine. These mutants were seen to have the greatest increase in specificity of all assayed mutants, including double mutants. Among the novel β-thioglycosidases addressed in this study, *Hor*GH1-M299R is the most promising mutant for the industrial application due to the larger turnover number.

AUTHOR CONTRIBUTION STATEMENT

FP conceived and designed research. NA conducted experiments. FP, NA, and NRM analysed data. NA and NRM drafted the manuscript and all authors read and approved the manuscript

ACKNOWLEDGEMENTS

The authors thank the King Faisal University in Saudi Arabia for financial support (NA). NRM and FP were funded by EraNet SUSFOOD2 (Project ID 122 ImPrOVE, through DEFRA UK) and by the Biotechnology and Biological Research Council [grant number BB/P002536/1].

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical standards

The authors conducted the research to the high ethical standards of the journal of submission.

REFERENCES

Babonneau D (2010) FitGISAXS: software package for modelling and analysis of GISAXS data using IGOR Pro. J Appl Cryst 43(4):929-936. doi:10.1107/S0021889810020352

Bourderioux A, Lefoix M, Gueyrard D, Tatibouet A, Cottaz S, Arzt S, Burmeister WP, Rollin P (2005) The glucosinolate-myrosinase system. New insights into enzyme-substrate interactions by use of simplified inhibitors. Org Biomol Chem. 3(10):1872-1879. doi:10.1039/b502990b

Burmeister WP, Cottaz S, Driguez H, Iori R, Palmieri S, Henrissat B (1997) The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl–enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. Phytochemistry 5(5):663–676. doi:10.1016/S0969-2126(97)00221-9

Chen, C., Natale, D.A., Finn, R.D., Huang, H., Zhang, J., Wu, C.H. and Mazumder, R (2011) Representative proteomes: a stable, scalable and unbiased proteome set for sequence analysis and functional annotation. Plos One 6(4). doi:10.1371/journal.pone.0018910

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: A Sequence Logo Generator. Genome Res 14(6). doi:10.1101/gr.849004

Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3:853-859. doi:10.1016/S0969-2126(01)00220-9

Dufour V, Stahl M, Baysse C (2015) The antibacterial properties of isothiocyanates. Microbiology 161(2):229-243. doi:10.1099/mic.0.082362-0

El-Gebali S, Mistry J, Batman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acid Res 47:D427-D432. doi:10.1093/nar/gky995

Gasteiger E, Hoogland C, Gattiker A, Duvaud Sv, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. in The proteomics protocols handbook 571-607. doi: 10.1385/1-59259-890-0:571

Halkier BA, Gershenzon J (2006) Biology and Biochemistry of Glucosinolate. Annu Rev Plant Biol 57:303-333. doi:10.1146/annurev.arplant.57.032905.105228

He XY, Wang XQ, Yang SJ, Chang WR, Liang DC (2001) Overexpression, purification, crystallization and preliminary crystallographic studies on a thermostable β -glycosidase from *Thermus nonproteolyticus* HG102. Acta Crystallogr D 57(11):1650-1651. doi:10.1107/s090744490101112x

Heckmann CM (2017) Construction of Expression Vectors for Selective His-tag Removal. University of Nottingham

Henn-Sax M, Höcker B, Wilmanns M, Sterner R (2001) Divergent evolution of (betaalpha)8-barrel enzymes. Biol Chem 382(9):1315-20. doi:10.1515/BC.2001.163

Husebye H, Arzt S, Burmeisterc WP, Frauke V. Hartelf, Brandtf A, Rossiterg JT, Bonesa AM (2005) Crystal structure at 1.1 Å resolution of an insect myrosinase from *Brevicoryne brassicae* shows its close relationship to β-glucosidases. Insect Biochem Molec 35(12):1311-1320. doi:10.1016/j.ibmb.2005.07.004

Jones AME, Winge P, Bones AM, Cole R, Rossiter JT (2002) Characterization and evolution of a myrosinase from the cabbage aphid *Brevicoryne brassicae*. Insect Biochem Molec 32(3):275-284. doi:10.1016/S0965-1748(01)00088-1

Kori LD, Hofmannb A, Patela BKC (2011) Expression, purification and preliminary crystallographic analysis of the recombinant β-glucosidase (BglA) from the halothermophile *Halothermothrix orenii*. Acta Crystallogr F 67(1):111-113. doi:10.1107/S1744309110046981

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Moleculr Evolutionary Genetics Analysis across Computing Platform. Mol Biol Evol 35(6) 1547-1549. doi:10.1093/molbev/msy096

Nong H, Zhang JM, Li DQ, Wang M, Sun XP, Zhu YJ, Meijer J, Wang QH (2010) Characterization of a novel β-thioglucosidase CpTGG1 in *Carica papaya* and its substrate-dependent and ascorbic acid-independent O-β-glucosidase activity. J Integr 52(10):879-890. doi:10.1111/j.1744-7909.2010.00988.x

Park S, Lee B, Park K (2017) Extremophilic Carbohydrate Active Enzymes (CAZymes). J Nutr Health Food Eng 7(1):1-9. doi:10.15406/jnhfe.2017.07.00230

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—A Visualization System for Exploratory Research and Analysis. J Comput Chem 25(13):1605-1612. doi:10.1002/jcc.20084

Pontoppidan B, Ekbom B, Eriksson S, Meijer J (2001) Purification and characterization of myrosinase from the cabbage aphid (*Brevicoryne brassicae*), a brassica herbivore. Euro J Biochem 268(4):1041-1048. doi:10.1046/j.1432-1327.2001.01971.x

Rakariyatham N, Butrindr B, Niamsup H, Shank L (2005) Screening of filamentous fungi for production of myrosinase. Braz J Microbiol 36(3): 242-245. doi:10.1590/S1517-83822005000300007

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Biochem Sci 16(6):276-277. doi:10.1016/s0168-9525(00)02024-2

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42(W1):W320-W324. doi:10.1093/nar/gku316

Samec D, Pavlovic´ I, Salopek-Sondi B (2017) White cabbage (Brassica oleracea var. capitata f. alba): botanical, phytochemical and pharmacological overview. Phytochem Rev 16(1):117-135. doi:10.1007/s11101-016-9454-4

Silverman JA, Balakrishnan R, Harbury PB (2001) Reverse engineering the (β/α)8 barrel fold. Proc Natl Acad Sci 98(6):3092-3097. doi:10.1073/pnas.041613598

Swift ML (1997) GraphPad Prism, Data Analysis, and Scientific Graphing. J Chem Inf Model 37(2):411-412. doi:10.1021/ci960402j

Wang X, He X, Yang S, An X, Chang W, Liang D (2003) Structural Basis for Thermostability of β -Glycosidase from the Thermophilic Eubacterium *Thermus nonproteolyticus* HG102. J Bacteriol Res 185(14):4248-4255. doi:10.1128/JB.185.14.4248–4255.2003

Winde I, Wittstock U (2011) Insect herbivore counteradaptations to the plant glucosinolate–myrosinase system. Phytochemistry 72(13):1566-1575. doi:10.1016/j.phytochem.2011.01.016

Witzel K, Hanschen FS, Klopsch R, Ruppel S, Schreiner M, Grosch R (2015) Verticillium longisporum infection induces organ-specific glucosinolate degradation in *Arabidopsis thaliana*. Front Plant Sci 6: 508. doi:10.3389/fpls.2015.00508

Yin J, Chen JC, Wu Q, Chen GQ (2015) Halophiles, coming stars for industrial biotechnology. Biotechnol Adv 33(7):1433-1442. doi:10.1016/j.biotechadv.2014.10.008

FIGURES

Fig. 1. Protein database structures of the active sites of glycoside enzymes. **a** *Thermus nonproteolyticus* glycoside hydrolase (pdb:1NP2); the residues mutated in this study are indicated, L171, H178 and V287. Residues essential for activity also indicated, E164 and E338. **b** *Halothermothrix orenii* glycoside hydrolase (pdb:3TA9); the residues mutated in this study are indicated, E173, H180 and M299. Residues essential for activity also indicated, E166 and E354. **c** diagram indicating the relationship between the substrate and the side chains of the residues of wild type *Thermus nonproteolyticus* mutated in this study, essential residues also displayed. Distances between the glycosidic substrate and sidechains predicted with UCSF Chimera. **d** diagram indicating the relationship between the glycosidic substrate and the side chains of the residues of wild type *Halothermothrix orenii* mutated in this study, essential residues also displayed. Distances between substrate and sidechains predicted with UCSF Chimera.

Fig. 2. Protein database structures of the active sites of myrosinase enzymes. **a** *S. alba* myrosinases (pdb:1E4M)**;** residues essential for activity also indicated, Q187 and E409. **b** *B. brassicae* myrosinases (pdb:1WCG); target residues of mutagenesis indicated, K173, Y180 and R312; residues essential for activity also indicated, E167 and E374. **c** diagram indicating the relationship between the thioglycosidic substrate and the side chains of the residues of wild type *B. brassicae* myrosinases targeted residues of mutagenesis, essential residues also displayed. Distances between substrate and sidechains predicted with UCSF Chimera.

Fig. 3. Amino acid sequence alignment of *Thermus nonproteolyticus* glycoside hydrolase (*T. nonproteolyticus)*, *Halothermothrix orenii* glycoside hydrolase (*H. orenii*), *B. brassicae* myrosinases (*B. brassicae*), *and S. alba* myrosinases (*S. alba*).

Fig. 4. Structures of the substrates, the native substrate of myrosinase is added to show the similarity in the structure of the molecules. **a** 4-nitrophenyl β -D-thioglucopyranoside (pNT-Glc). **b** 4-nitrophenyl- β -D-Glucopyranoside (pNP-Glc). **c** sinigrin, the native substrate of myrosinase.

Fig. 5 Sequence logo generated for amino acid position 178 (*Thermus nonproteolyticus* glycoside hydrolase numbering) of glycosyl hydrolase family 1, indicating that cystine and histidine are the most highly conserved at position 178. Sequences were taken from Pfam, the representative proteome at 15% co-membership threshold (approximately 3900 sequences) was aligned with the wild type sequences of the β-glycosidase used in this study. Sequence logo was generated with WebLogo.

Table 1: Table summarizing the kinetic parameters of the wild type enzyme and mutants of *Thermus nonproteolyticus* (*Tno*GH1) against the 4-nitrophenyl β -D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl- β -D-Glucopyranoside (pNP-Glc). Experiments were conducted in triplicate. Standard errors are given, based on fitted theoretical curves.

	k_{cat}/K_M (pNT-Glc) / $M^{-1}s^{-1}$	k_{cat} (pNT-Glc)/ s^{-1}	K_M(pNT-Glc)/ mM	k_{cat}/K_M(pNP-Glc)/ $M^{-1}s^{-1}$	k_{cat}(pNP-Glc)/ s^{-1}	K_M(pNP-Glc)/ mM
WT *Tno*GH1	940 ± 70	1.34 ± 0.04	1.40 ± 0.10	$246 \times 10^3 \pm 3 \times 10^3$	132 ± 3	0.54 ± 0.06
*Tno*GH1-L171K	1200 ± 100	1.10 ± 0.04	0.90 ± 0.10	$92 \times 10^3 \pm 9 \times 10^3$	136 ± 4	1.50 ± 0.20
*Tno*GH1-V287R	2500 ± 400	1.39 ± 0.01	0.56 ± 0.01	$190 \times 10^3 \pm 20 \times 10^3$	117 ± 3	0.61 ± 0.07
*Tno*GH1-H178Y	300 ± 50	0.15 ± 0.01	0.27 ± 0.07	$230 \times 10^3 \pm 36 \times 10^3$	77 ± 2	0.34 ± 0.06
*Tno*GH1-L171K/V287R	1820 ± 320	0.08 ± 0.01	0.21 ± 0.04	$58 \times 10^3 \pm 5 \times 10^3$	116 ± 3	1.90 ± 0.20
*Tno*GH1-L171K/H178Y	380 ± 80	0.09 ± 0.04	0.22 ± 0.05	$110 \times 10^3 \pm 10 \times 10^3$	69 ± 2	0.70 ± 0.10
*Tno*GH1-V287R/H178Y	2420 ± 500	0.84 ± 0.04	0.35 ± 0.08	$260 \times 10^3 \pm 40 \times 10^3$	106 ± 3	0.41 ± 0.07

Table 2: Table summarizing the kinetic parameters of the wild type enzyme and mutants of *Halothermothrix orenii* (*Hor*GH1) against the 4-nitrophenyl β -D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc). Experiments were conducted in triplicate. Standard errors are given, based on fitted theoretical curves.

	k_{cat}/K_M(pNT-Glc) / $M^{-1}s^{-1}$	k_{cat} (pNT-Glc)/ s^{-1}	K_M(pNT-Glc)/ mM	k_{cat}/K_M(pNP-Glc)/ $M^{-1}s^{-1}$	k_{cat}(pNP-Glc)/ s^{-1}	K_M(pNP-Glc)/ mM
WT *Hor*GH1	9083 ± 529	37.2 ± 0.7	4.1 ± 0.3	102000 ± 7 × 10^3	80 ± 1	0.5 ± 0.1
*Hor*GH1-E173K	3331 ± 153	19.1 ± 0.3	5.7 ± 0.3	11000 ± 1 × 10^3	26 ± 2	2.5 ± 0.5
*Hor*GH1-M299R	13260 ± 170	33.7 ± 1.2	2.5 ± 0.4	36000 ± 2 × 10^3	34 ± 1	0.9 ± 0.1
*Hor*GH1-H180Y	2501 ± 131	9.4 ± 0.2	3.8± 0.2	55000 ± 5 × 10^3	38 ± 1	0.7 ± 0.1
*Hor*GH1-E173K/M299R	2622 ± 122	10.1 ± 0.2	3.8 ± 0.2	7000 ± 1 × 10^3	12 ± 1	1.6 ± 0.1
*Hor*GH1-E173K/H180Y	1080 ± 61	5.2 ± 0.1	4.8 ± 0.3	15000 ± 1 × 10^3	24 ± 1	1.8 ± 0.2
*Hor*GH1-M299R/H180Y	3770 ± 198	7.3 ± 0.1	1.9 ± 0.1	23000 ± 3 × 10^3	25 ± 2	1.1 ± 0.2